Global Interbull EBV in domestic single step genomic evaluation

  J. Bauer,  J. Přibyl,  J. Přibylová,  
  L. Vostrý,  L. Zavadilová,

Czech Moravian Breeding Corporation, Praha , Czech Rep.  
  V. Čermák,  Z. Růžička,  J. Šplíchal,  M. Verner,

Holstein Cattle Breeders Association of the Czech Rep.  
  J. Motyčka,  L. Vondrášek

* vostry@af.czu.cz

supported MZe ČR
Objective:

Combine for Holstein cattle domestic database of production records into one joint evaluation.

Interbull EBV
Circumstances

Majority of Sires are foreign

Therefore sires in insemination weak connection to the domestic population.

But
criterion is the production in domestic condition.
SNP editing:

- MAF,
- G-score,
- No. of loci per bull,
- No. of bulls per locus,
- Big error of prediction of old bulls in training set,
- Big discrepancy of relationship $A_{22} \times G$,
- Proportion of H.
After editing

- 39,904 loci
  - 1,605 genotyped bulls,
    - 1,259 in training set,
      - 140 with 0 daughters in 2008 and > 50 (average 67) in 2012,
    - 206 others, with small No. of daughters.
Data
1,958,139 animals in pedigree

1,185,225 first lactations II.

969,269 first lactations I.

1,762,905 animals in pedigree

1,605 genotyped bulls

2012

Bulls with daughters’ production 1,259

> 50 daughters with production

1991

1,259 young bulls 140
Domestic records

Interbull 98,037 EBVs

Genotyped bulls
deregressed DRP, ERC

converted to EBV CZ
## Strategies

<table>
<thead>
<tr>
<th>Method of prediction</th>
<th>Calculated value</th>
<th>Sources of production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Domestic</td>
</tr>
<tr>
<td>BLUP</td>
<td>EBV</td>
<td>D-EBV</td>
</tr>
<tr>
<td>Ridge Regression</td>
<td>DGV</td>
<td>rl-DGV</td>
</tr>
<tr>
<td></td>
<td>GEBV *</td>
<td>rl-GEBV</td>
</tr>
<tr>
<td>GBLUP</td>
<td>DGV</td>
<td>gl-DGV</td>
</tr>
<tr>
<td></td>
<td>GEBV *</td>
<td>gl-GEBV</td>
</tr>
<tr>
<td>ssGBLUP **</td>
<td>GEBV</td>
<td>D-GBEBV</td>
</tr>
</tbody>
</table>

* GEBV = 0.8 DGV + 0.2 D-EBV.

** Weight for genomic relationship $G$ 80 % and pedigree relationship $A22$ 20 %.
# Size of data for prediction 2008

<table>
<thead>
<tr>
<th>Source</th>
<th>Records</th>
<th>Weights ERC</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td>969,269</td>
<td>969,269</td>
<td>D- EBV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-GEBV</td>
</tr>
<tr>
<td>Interbull for genotyped bulls</td>
<td>1,259</td>
<td>57,864</td>
<td>rl- DGV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rl-GEBV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>gl- DGV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>gl-GEBV</td>
</tr>
<tr>
<td>Interbull for all bulls</td>
<td>98,037</td>
<td>785,276</td>
<td>I- EBV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I-GEBV</td>
</tr>
<tr>
<td>D + I all</td>
<td>1,064,912</td>
<td>1,632,668</td>
<td>DI- EBV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI-GEBV</td>
</tr>
</tbody>
</table>
Weighted analysis

Ridge Regression

\[ y_j = \mu + \Sigma \delta_i \cdot g_i + e_j \]

Fixed effects

- \( \mu \) – contemporary group

Random effects

- \( g_i \) – locus
- \( \delta_i \) – regression coefficients
- \( e_j \) – residual

Heritability \((h^2) = 0.25\); Weight = ERC
Weighted analysis

GBLUP

\[ y_j = \mu + a_{nj} + e_j \]

**Fixed effects**
- \( \mu \) – contemporary group

**Random effects**
- \( a_{nj} \) – animal
- \( e_j \) – residual

Heritability \((h^2) = 0.25\); Weight = ERC
**Weighted analysis**

**BLUP / ssGBLUP**

\[ y_{ijkl} = HYS_i + \beta_1 \cdot ca_j + \beta_2 \cdot ca_j^2 + \beta_3 \cdot do_k + \beta_4 \cdot do_k^2 + an_l + e_{ijkl} \]

**Fixed effects**
- **HYS** – contemporary group
- **\( \beta \)** – regression coefficients
- **ca_j; ca_j^2** – curvilinear regression on calving age
- **do_k; do_k^2** – curvilinear regression on days open

**Random effects**
- **an_l** – animal
- **e_{ijkl}** – residual

Heritability (\( h^2 \)) = 0.25; Weight = ERC
Relationship

Ridge Regression...I
BLUP – AM ............A
GBLUP....................G (VanRaden 2008)
ssGBLUP....................H (Legarra et al., 2009)

20% A_{22}  G  80%

\( \mathbf{G} \) – normalised (aver. diag. = 1)

(shifted (aver. \( \mathbf{G} = \text{aver.} \mathbf{A}_{22} \))

(Forni et al., 2011)

(Vitezica et al., 2011)
Programs

• BLUPF90 family of programs
  (Misztal et al., 2002)

• DMU5 module of the DMU software
  (Madsen & Jensen, 2010)

• G-matrix
  (Su & Madsen, 2011)
Verification Procedure

• For the 140 young bulls:

  Correlation:
  EBV/GEBV/DGV from Data I (2008) ×
  EBV/DYD from Data II (2012)

• Data I – using only pedigree information and no progeny information
• Data II – after progeny testing.
Results
Difference of young bulls from basis of 2,116 sires with > 60 daughters in 2008

<table>
<thead>
<tr>
<th>Data 2008</th>
<th>Mean milk kg</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td>657</td>
<td>D- EBV</td>
</tr>
<tr>
<td></td>
<td>672</td>
<td>D-GEBV</td>
</tr>
<tr>
<td>D + I all</td>
<td>651</td>
<td>DI- EBV</td>
</tr>
<tr>
<td></td>
<td>640</td>
<td>DI-GEBV</td>
</tr>
<tr>
<td>Data 2012</td>
<td>629</td>
<td>D- EBV</td>
</tr>
<tr>
<td></td>
<td>630</td>
<td>D-GEBV</td>
</tr>
</tbody>
</table>
## Correlation Interbull data 2008

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>rl- DGV</td>
<td>0.60</td>
<td>0.57</td>
</tr>
<tr>
<td>rl- GEBV</td>
<td>0.67</td>
<td>0.61</td>
</tr>
<tr>
<td>gl- DGV</td>
<td>0.59</td>
<td>0.57</td>
</tr>
<tr>
<td>gl-GEBV</td>
<td>0.66</td>
<td>0.61</td>
</tr>
<tr>
<td>l- EBV</td>
<td>0.62</td>
<td>0.53</td>
</tr>
<tr>
<td>l-GEBV</td>
<td><strong>0.70</strong></td>
<td><strong>0.65</strong></td>
</tr>
</tbody>
</table>
Correlation Domestic + Interbull data 2008

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D- EBV</td>
<td>0.59</td>
<td>0.47</td>
</tr>
<tr>
<td>D-GEBV</td>
<td>0.70</td>
<td>0.61</td>
</tr>
<tr>
<td>DI- EBV</td>
<td>0.63</td>
<td>0.51</td>
</tr>
<tr>
<td>DI-GEBV</td>
<td><strong>0.73</strong></td>
<td><strong>0.64</strong></td>
</tr>
</tbody>
</table>
Validated reliability
Conclusions

• Combination of domestic and Interbull database improves the prediction of both EBV and GEBV, especially in small populations.

• The DI-GEBV by ssGBLUP was the most suitable method.
Acknowledgments

We are grateful to Dr. Per Madsen, University of Aarhus, Denmark, and to Prof. Ignacy Misztal, University of Georgia, USA, for methodical support and for provision of computer programs DMU.